AD-A037 766 INDIANA UNIV BLOOMINGTON DEPT OF MATHEMATICS F/G 12/1
CONVERGENCE AND REMAINDER TERMS IN LINEAR RANK STATISTICS.(U)
JAN 77 H BERGSTROM, M L PURI AF-AFOSR-2927-76
UNCLASSIFIED AFOSR-TR-77-0168 NL

1 OF 1
AD
A037766

END
DATE
FILMED
4-77

UNCLASSIFIED

(4) NW

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFOSR - TR - 77 - 0168	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): CONVERGENCE AND REMAINDER TERMS IN LINEAR RANK STATISTICS.		5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Harold Bergstrom and Madan L. Puri		8. CONTRACT OR GRANT NUMBER(s): AF-AFOSR-2927-76
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Indiana University, Department of Mathematics, Bloomington, Indiana 47401 (12) 23p.		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332		12. REPORT DATE: Janaury 1977
		13. NUMBER OF PAGES: 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (11) Jan 77		15. SECURITY CLASS. (of this report): UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

COPY AVAILABLE TO DDC DOES NOT PERMIT FULLY LEGIBLE PRODUCTION

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

linear rank statistics, score generating functions, rate of convergence.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A new approach to the asymptotic normality of simple linear rank statistics for the regression case, studied earlier by Hajek (1968; Ann. Math. Statist, 39, 325-346) is provided along with the estimation of the remainder terms in the approximation to normality.

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 402 522

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

4

Convergence and Remainder Terms In Linear Rank Statistics

by

Harold Bergström and Madan L. Puri
University of Göteborg, Chalmers University of Technology, and Indiana University.

A new approach to the asymptotic normality of simple linear rank statistics for the regression case studied earlier by Hájek (1968) is provided along with the estimation of the remainder term in the approximation to normality.

1. Introduction and Summary. Let $X_1,\ldots,X_n$ be independent random variables having continuous cdfs (cumulative distribution functions) $F_1(x),\ldots,F_n(x)$ respectively. Consider a statistic $S_n = s(X_1,\ldots,X_n)$ with $ES_n = 0$ and

---

Work supported by the Air Force Office of Scientific Research, AFSC, USAF, under Grant No. AFOSR 76-2927 . Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

AMS 1970 subject classifications. Primary 62E20 ; Secondary 60F05, 60F99 Key words and phrases. Linear rank statistics, score generating functions, rate of convergence.

$ES_n^2 < \infty$. Then, to prove the asymptotic normality of S_n (as $n \to \infty$) , Hájek (1968) uses the method of projection which gives to the statistic S_n , the approximation of the form

$$\hat{S}_n = \sum_{j=1}^{n} E[\, S_n | X_j \,] \quad . \tag{1.1}$$

Consider now the simple linear rank statistic S_n introduced by Hájek (1962, 1968)

$$S_n = \sum_{j=1}^{n} c_j \{ \psi(R_j/n) - E[\psi(R_j/n)] \} \tag{1.2}$$

where the c's are known constants, R_j is the rank of X_j among $(X_1, \ldots, X_n)$ and $\psi(\cdot)$ is a score generating function defined on $(0,1)$. Hájek (1962) [see also Hájek-Šidák (1967)] established the asymptotic normality of S_n in (1.2) under the assumption that the F_i are contiguous, e.g. when $F_i(x) = F(x - \Delta\, d_{ni})$ where Δ is the unknown parameter and the d's are the known constants. Later on Hájek (1968) studied the asymptotic normality of S_n for the general $F_i(x)$ (the non-contiguous case). Under the set-up of Hájek (1962), Jurečková and Puri (1975) , referred to hereafter as JP, studied the problem of determining the rate of convergence of the cdf of S_n to the limiting normal cdf and established it of order $O(N^{-\frac{1}{2}+\delta})$ for $\delta > 0$. In this paper we not only give a new approach to the asymptotic normality of S_n for the general F_i (i.e. not necessarily contiguous) but improve the results of JP in providing a sharper bound (for the general F_i's). In the passing, we may also mention that where as JP requires ψ to have a bounded fourth derivative, here we only require the boundedness of the second derivative. Furthermore whereas this paper gives more explicit error bounds than the JP paper, the later gives more information on the limiting behavior of ES_n and $\mathrm{Var}\, S_n$.

We now introduce some notations. We define $\psi(\cdot) = 0$ outside $(0,1)$. Then, we can use the supremum norm

$$\|\psi\| = \sup_{t\in(-\infty,\infty)} |\psi(t)| \tag{1.3}$$

Set

$$\rho_i = R_i/n\,,\ \rho_{ii} = E[\rho_i | X_i]\,,\ u(x) = 1 \text{ if } x \geq 0 \text{ and } u(x) = 0\,, \text{ otherwise.} \tag{1.4}$$

Then

$$R_i = \sum_{j=1}^{n} u(X_i - X_j)\,. \tag{1.5}$$

In this paper, we shall deal with the following approximation of S_n.

$$(1.6) \qquad T_n = \sum_{i=1}^{n} c_i\{\psi(\rho_{ii}) - E[\psi(\rho_{ii})] + (\rho_i - \rho_{ii})\psi'(\rho_{ii})\} \ ,$$

assuming that ψ' exists on $(0,1)$ and

$$(1.7) \qquad \hat{T}_n = \sum_{j=1}^{n} E[T_n | X_j] \ .$$

Since $E[(\rho_i - \rho_{ii})\psi'(\rho_{ii})] = 0$, it follows that

$$(1.8) \quad \hat{T}_n = \sum_{i=1}^{n} c_i\{\psi(\rho_{ii}) - E[\psi(\rho_{ii})] + \sum_{j\neq i}^{n} E[(\rho_i - \rho_{ii})\psi'(\rho_{ii}) | X_j]\} \ .$$

Let H_n , G_n and $\hat{G}_n$ be the cdfs of S_n , T_n and $\hat{T}_n$ respectively, and put

$$(1.9) \quad \sigma_n^2 = E[S_n^2] \ , \quad \hat{\delta}_n^2 = E[\hat{T}_n^2] \ , \quad \Gamma_{nr}^{2r} = \frac{1}{n}\sum_{i=1}^{n} c_i^{2r} \ , \ \Gamma_{nr} > 0 \ .$$

Then our theorems are the following:

<u>Theorem 1.1</u> : <u>If</u> ψ <u>has a derivative on</u> $(0,1)$ <u>then</u>

$$(1.10) \ \|\hat{G}_n(\hat{\delta}_n) - \Phi(\cdot)\| \leq 4C[2\|\psi\|^3 + \|\psi'\|^3] \sum_{i=1}^{n} |c_i|^3 \hat{\delta}_n^{-3} \ ;$$

$$\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

<u>where</u> C <u>is the constant in Berry-Esseen's inequality</u> (Zolotarev (1967) gives the approximation 0,9051) . Further

$$(1.11) \qquad |\hat{\delta}_n - \sigma_n| \leq c_1(\|\psi'\| + \|\psi''\|)\Gamma_{n,1}$$

with an absolute constant C_1 , provided ψ'' exists on $(0,1)$.

Theorem 1.2 : If ψ has a second order derivative on $(0,1)$, then for any positive integers n and r such that $n \geq 2r$

$$(1.12) \qquad \|H_n(\hat{\delta}_n \cdot) - \Phi(\cdot)\| \leq 4C(2\|\psi\|^3 + \|\psi'\|^3) \sum_{i=1}^{n} |c_i|^3 \hat{\delta}_n^{-3}$$

$$+ C_2[\hat{\delta}_n^{-1}(\|\psi'\| + \|\psi''\|) r\Gamma_{nr}]^{\frac{2r}{2r+1}} \quad .$$

where C_2 is an absolute constant.

Remark. If the c_i are choosen such that $|c_i| \leq a/\sqrt{n}$ with constant a for all i and n , then

$$\Gamma_{nr} \leq a/\sqrt{n} \quad ,$$

and for $r = [\log n]$, $[r\Gamma_{nr}]^{\frac{2r}{2r+1}} \leq a \sqrt[4]{e} (\log n)\, n^{-\frac{1}{2}}(1+0(\frac{1}{\log n}))$

Note that $\hat{\delta}_n^{-1} c_i$ is invariant and thus also $\hat{\delta}_n^{-1} \Gamma_{nr}$ is invariant under the transformation $c_i \to \gamma c_i$, $i = 1,2,\dots$.

2. Same Lemmas.

Lemma 2.1 : For any positive integers r and n , $2r \leq n$, we have

$$(2.1) \qquad E[(\rho_i - \rho_{ii})^{2r}] \leq b(r) n^{-r}$$

with

$$(2.2) \qquad b(r) \le n^{-r} \sum_{t=1}^{r} \binom{n-1}{t} \frac{(2r)!}{(2r-2t)!} t^{2r-2t} \cdot 2^{-3t}$$

and for $n^{-1} r^3 \le \frac{3}{4}$

$$(2.3) \qquad b(r) \le 2^{-3r} \frac{(2r)!}{r!} [1 + 8n^{-1} r^3]$$

Proof: By (1.4) we obtain

$$\rho_i - \rho_{ii} = \frac{1}{n} \sum_{j \ne i}^{n} [u(X_i - X_j) F_j(X_i)] \; .$$

By the polynomial theorem we then get

$$(2.4) \quad E[(\rho_i - \rho_{ii})^{2r}] = n^{-2r} \sum_{s_1 + \ldots + s_n = 2r} \frac{(2r)!}{s_1! \ldots s_n!} E \prod_{j \ne 1}^{n} [u(X_i - X_j) - F_j(X_i)]^{s_j} \; .$$

We claim that any term in this sum is equal to zero if $s_{j_0} = 1$ for some j_0. Indeed we find that the conditional expection of the product with respect to all X_j, $j \ne j_0$ is equal to 0 if $s_{j_0} = 1$. Hence we have only to regard terms with $s_j = 0$ or ≥ 2 for any j, and there can be at most $t \le r$ exponents s_j different from 0. If $s_j \ge 2$, $j = 1,2,\ldots t$, $s_j = 0$ for $j > t, i > t$ we obtain, observing that

$$|u(X_i - X_j) - F_j(X_i)| \leq 1$$

$$(2.5) \qquad E[\prod_{j=1}^{t} [u(X_i - X_j) - F_j(X_i)]^{s_j} \leq E \prod_{j=1}^{t} [u(X_i - X_j) - F_j(X_i)]^2$$

$$= E[\prod_{j=1}^{t} [F_j(X_i) - F_j^2(X_i)]] \leq 4^{-t} \quad .$$

This inequality remains true for all permutations of the indices $1,\ldots,n$. Put

$$(2.6) \qquad \gamma(t) = \sum_{\substack{s_1+\ldots+s_t = 2r \\ s_j \geq 2, j=1,\ldots,t}} \frac{(2r)!}{s_1!\ldots s_t!} \quad .$$

Since t indices out of $n-1$ indices can be choosen in $\binom{n-1}{t}$ different ways we obtain from (2.4) through (2.6),

$$(2.7) \qquad E[(\rho_i - \rho_{ii})^{2r}] \leq n^{-2r} \sum_{t=1}^{r} \binom{n-1}{t} \gamma(t) 4^{-t} \quad .$$

We claim that

$$(2.8) \qquad \gamma(t) \leq \frac{(2r)!}{(2r-2t)!} 2^{-t} t^{2r-2t} \quad .$$

Indeed, differentiating the identity

$$\left(\sum_{j=1}^{t} y_j\right)^{2r} = \sum_{s_1+\ldots+s_t=2r} \frac{(2r)!}{s_1!\ldots s_t!} \prod_{j=1}^{t} y_j^{s_j}$$

twice with respect to all y_j and then putting all y_j equal to 1, we obtain

$$\frac{(2r)!}{(2r-2t)!} t^{(2r-2t)} = \sum_{\substack{s_1+\ldots+s_t = 2r \\ s_j \geq 2,\ j = 1\ldots t}} \prod_{j=1}^{t} s_j(s_j-1) \frac{(2r)!}{s_1!\ldots s_t!}$$

Now using (2.7) and (2.8), we get (2.1) and (2.2). We now estimate $b(r)$ further, mainly for use when n and r are large. Put $r-t = u$. Then we can write

$$(2.9) \qquad b(r) \leq 2^{-3r} \sum_{u=0}^{r-1} k(u)$$

with

$$k(u) = \frac{n^{-u}(2r)!\ (r-u)^{2u} 2^{3u}}{(r-u)!\ (2u)!} .$$

Particularly

$$k(0) = \frac{(2r)!}{r!},\ k(1) < 4n^{-1}r^3 \cdot \frac{(2r)!}{r!}$$

and for $u \geq 1$

$$\frac{k(u+1)}{k(u)} = n^{-1} \left(1 - \frac{1}{r-u}\right)^{2u} \cdot 2^3 \cdot (r-u) \frac{(r-u-1)^2}{(2u+1)(2u+2)}$$

$$< \frac{2}{3} n^{-1} r^3 \leq \tfrac{1}{2} \quad \text{for} \quad n^{-1}r^3 \leq \frac{3}{4} .$$

Hence

$$b(r) \le 2^{-3r} \cdot \frac{(2r)!}{r!} [1 + 8n^{-1}r^3]$$

for $n^{-1}r^3 \le \frac{3}{4}$.

Lemma 2.2 : For any positive integers r and n, $2r \le n$, we have

$$E(T_n - \hat{T}_n)^{2r} \le c(r) \|\psi'\|^{2r} \Gamma_{n,r}^{2r} \tag{2.10}$$

if ψ' exists on $(0,1)$, and if ψ'' exists on $(0,1)$

$$E[(S_n - T_n)^{2r}] \le b(2r) \|\psi''\|^{2r} \Gamma_{n,r}^{2r} , \tag{2.11}$$

$$E[(S_n - \hat{T}_n)^{2r}] \le d(r,\psi) \Gamma_{n,r}^{2r} \tag{2.12}$$

with

$$b(2r) \le n^{-2r} \sum_{t=1}^{2r} \binom{n-1}{t} \frac{(4r)!}{(4r-2t)!} t^{4r-2t} \cdot 2^{-3t}$$

$$c(r) \le 2^{2r} n^{-2r} \sum_{t=1}^{2r} \binom{n}{t} \frac{(4r)!}{(4r-2t)!} t^{4r-2t} \cdot 2^{-t}$$

$$d(r,\psi) \le [[b(2r)]^{\frac{1}{2r}} \|\psi''\| + [c(r)]^{\frac{1}{2r}} \|\psi'\|]^{2r} .$$

Further we have the estimates

$$(2.13) \qquad b(2r) \le 2^{-6r}\,\frac{(4r)!}{(2r)!}\,[1+2^6 n^{-1} r^3]$$

for $2^3 n^{-1} r^3 \le \frac{3}{4}$,

$$(2.14) \qquad c(r) \le \frac{(4r)!}{(2r)!}[1+2^3 n^{-1} r^3] \quad \text{for} \quad n^{-1} r^3 \le \frac{3}{8}$$

<u>Remark</u> : <u>By Stirling's approximation of the Γ-function we have</u>

$$\frac{(4r)!}{(2r)!} \le 2^{6r+\frac{1}{2}}\, r^{2r}\,(\exp -2r)\,\exp\frac{1}{48r} \;.$$

<u>Proof</u> : By (1.6) and (1.8) we get

$$(2.15) \qquad T_n-\hat{T}_n = \sum_{i=1}^{n} c_i\{(\rho_i-\rho_{ii})\psi'(\rho_{ii}) - \sum_{\substack{j=1\\ j\neq i}}^{n} E[(\rho_i-\rho_{ii})\psi'(\rho_{ii})\,|\,X_j]\}$$

and for $j \neq i$

$$(2.16) \qquad E[(\rho_i-\rho_{ii})\psi'(\rho_{ii})\,|\,X_j] = \frac{1}{n}\sum_{k\neq i}^{n} E\{[u(X_i-X_k) -$$

$$- F_k(X_i)]\psi'(\rho_{ii})\,|\,X_j\} = \frac{1}{n}E[u(X_i-X_j)-F_j(X_i)]\psi'(\rho_{ii})\,|\,X_j]$$

since the conditional expectations in the sum are zero for $j \neq k, i$. Now using the relation

$$(\rho_i-\rho_{ii})\psi'(\rho_{ii}) = \frac{1}{n}\sum_{j\neq i}^{n}[u(X_i-X_j) - F_j(X_i)]\psi'(\rho_{ii})$$

and noting that

$$E[(\rho_i-\rho_{ii})\psi'(\rho_{ii})|X_i] = 0$$

we obtain from (2.15)

(2.17) $$T_n - \hat{T}_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n} c_i V_{ij}$$

with

(2.18) $$V_{ij} = [u(X_i-X_j) - F_j(X_i)]\psi'(\rho_{ii}) - E\{[u(X_i-X_j) - F_j(X_i)]\psi'(\rho_{ii})|X_j\} .$$

Clearly

(2.19) $$E[V_{ij}|X_j] = 0 , \quad E[V_{ij}|X_i] = 0 .$$

By the polynomial theorem we get

(2.20) $$E[(T_n-\hat{T}_n)^{2r}] = n^{-2r}E[\sum_{i=1}^{n}\sum_{j\neq i}^{n} c_i V_{ij}]^{2r}$$

$$= n^{-2r}\sum \frac{(2r)!}{\prod_{i=1}^{n}\prod_{j\neq i}^{n}(s_{ij}!)} E\{\prod_{i=1}^{n}\prod_{j\neq i}^{n}(c_i V_{ij})^{s_{ij}}\}$$

where the sum should be taken over terms corresponding to different vector solutions $\{s_{ij}\}$, i, j = 1,...n , $j \neq i$

of the equation

$$(2.21) \qquad \sum_{i=1}^{n} \sum_{j\neq i}^{n} s_{ij} = 2r \ .$$

The expectation

$$(2.22) \qquad E[\prod_{i=1}^{n} \prod_{j\neq i}^{n} V_{ij}^{s_{ij}}]$$

is equal to 0 for some vector solutions of (2.21) since (2.19) holds, and we have only to regard those solutions for which the expectation (2.22) is not equal to 0 .

We say that s_{ij} gives the contribution $\frac{1}{2} s_{ij}$ to the sum (2.21) from each of the indices i and j . Hence according to this notation an index k gives the contribution

$$(2.23) \qquad g(k) = \frac{1}{2} \sum_{j\neq k}^{n} s_{kj} + \frac{1}{2} \sum_{j\neq k}^{n} s_{jk}$$

to the sum (2.21). By conditioning with respect to all X_j , $j \neq k$ we easily find that the expectation (2.22) is equal to 0 if k gives the contribution $\frac{1}{2}$ to the sum (2.21), i.e. if $s_{kj} = 1$ for exactly one index $j \neq k$, and $s_{jk} = 0$ for $j \neq k$ or if $s_{jk} = 1$ for exactly one j and $s_{kj} = 0$ for $j \neq k$.

The sum Σ on the right hand side of (2.20) can be divided into partial sums as follows. Let C be a collection of different positive integers belonging to the set

$1,\ldots,2r$, say $C = (1,2,\ldots,t)$. Let Σ_C consist of all terms in (2.20) corresponding to the vector solutions of (2.21) such that

(a) $s_{ij} = 0$ if not both i and j belong to C ;

(b) for any $k \in C$ the contribution to the sum (2.21) is larger than 1/2 . Note that C can contain at most $2r$ different integers since every $k \in C$ gives at least the contribution 1 to the sum (2.21). Clearly partial sums Σ_{C_1} and Σ_{C_2} contain no common terms if $C_1 \neq C_2$. Consider now the expectation

$$E\Big[\prod_{i=1}^{t}\prod_{j\neq i}^{t}(c_i V_{ij})^{s_{ij}}\Big]$$

where the i and j belong to the collection C . Note that s_{ij} may be equal to 0 for some pairs (i,j) . By Hölder's inequality we get, using the fact that $|V_{ij}| \leq 2\|\psi'\|$

$$\text{(2.24)}\qquad \Big|E\prod_{i=1}^{t}\prod_{j\neq i}^{t}(c_i V_{ij})^{s_{ij}}\Big| \leq \prod_{i=1}^{t}\prod_{j\neq i}^{t}|c_i|^{s_{ij}}\{E[(V_{ij})^{2r}]\}^{\frac{s_{ij}}{2r}}$$

$$\leq 2^{2r}\|\psi'\|^{2r}\prod_{i=1}^{t}|c_i|^{s_i}$$

where

$$\text{(2.25)}\qquad s_i = \sum_{j=1}^{t} s_{ij}\ , \quad \sum_{i=1}^{t} s_i = 2r\ .$$

The partial sum corresponding to C is then estimated by

$$(2.26)\qquad \Sigma_C' \frac{(2r)!}{\prod_{i=1}^{t} \prod_{i\neq j}^{t} (s_{ij})!} \left(2^{2r}\|\psi'\|^{2r} \prod_{i=1}^{t} |c_i|^{s_i}\right) .$$

Note that $\frac{(2r)!}{\prod_{i=1}^{t} \prod_{i\neq j}^{t} (s_{ij}!)}$ is an integer. Hence

we have $N(t) = \Sigma_C' \frac{(2r)!}{\prod_{i=1}^{t} \prod_{j\neq i}^{t} (s_{ij})!}$

terms in the class C which are estimated by (2.24) . Let C_t be the set of all terms

$$\Sigma E \prod_{i=1}^{n} \prod_{j\neq 1}^{n} (c_i V_{ij})^{s_{ij}}$$

in (2.2D) which belong to some class C containing exactly t indices. Let $(s_1, s_2, \ldots, s_t)$ in (2.26) be given, $0 \leq s_1 \leq s_2 < \ldots \leq s_t$, $\sum_{i=1}^{t} s_i = 2r$. Then according to the symmetry the set C_t contains a sum of terms, each estimated by

$$(2.27)\qquad 2^{2r}\|\psi'\|^{2r} \prod_{i=1}^{t} |c_{k_i}|^{s_i}$$

where $(k_1 \ldots k_t)$ is any combination of numbers 1, 2, ..., n to the t^{th} class and in any order within this class. Let the number of terms in C_t for a fixed vector

$(s_1,s_2,\dots,s_t)$ as above be $n(t)$ and the sum of terms (2.27) belonging to $(s_1,s_2,\dots,s_t)$ be $A(s_1,s_2,\dots,s_t)$. (Note that $n(t)$ depends on $s_1,\dots,s_t$). Then, since $A(s_1,\dots,s_t)$ is a symmetrical function

$$(2.28)\quad A(s_1,s_2,\dots,s_t) = \frac{n(t)}{n!}\,\Sigma' 2^{2r}\|\psi'\|^{2r}\prod_{i=1}^{t}|c_{k_i}|^{s_i}$$

where Σ' is the sum all terms belonging to all permutations of the numbers $1,2,\dots n$. By Hölder's inequality we get observing that

$$|c_{k_i}|^{s_i} = [c_{k_i}^{2r}]^{\frac{s_i}{2r}}, \quad \sum_{i=1}^{t}\frac{s_i}{2r} = 1,$$

$$(2.29)\quad \Sigma'\prod_{i=1}^{t}|c_{k_i}|^{s_i} \le \prod_{i=1}^{t}(\Sigma' c_{k_i}^{2r})^{\frac{s_i}{2r}}$$

and here

$$\Sigma' c_{k_i}^{2r} = \frac{n!}{n}\sum_{i=1}^{n} c_i^{2r} .$$

Hence we obtain by (2.28) and (2.29)

$$A(s_1,s_2,\dots,s_t) \le 2^{2r}\|\psi'\|^{2r}\cdot n(t)\cdot\frac{1}{n}\sum_{i=1}^{n} c_i^{2r} .$$

Since C_t contains $\binom{n}{t}N(t)$ terms we then find that C_t gives at most the contribution

$$n^{-2r}2^{2r}\|\psi'\|^{2r}\binom{n}{t}N(t)\cdot\frac{1}{n}\sum_{i=1}^{n}c_i^{2r}$$

to the right hand side of (2.20). Putting

$$\Gamma_{nr}^{2r}=\frac{1}{n}\sum_{i=1}^{n}c_i^{2r}\ ,\quad \Gamma_{nr}\geq 0\ ,$$

and regarding the sets $\mathcal{C}_t$ for $t = 1,2,\ldots,2r$, we obtain from (2.20) that

$$\text{(2.30)}\qquad E[(T_n-\hat{T}_n)^{2r}]\leq 2^{2r}n^{-2r}\|\psi'\|^{2r}\Gamma_{nr}^{2r}\cdot\sum_{t=1}^{2r}\binom{n}{t}N(t)\ .$$

We estimate $N(t)$ in the following way. Consider the identity

$$\text{(2.31)}\qquad \Big(\sum_{i=1}^{t}\sum_{j=i}^{t}x_ix_j\Big)^{2r}=\sum\frac{(2r)!}{\prod_{i=1}^{t}\prod_{j=1}^{t}(s_{ij})!}\prod_{i=1}^{t}\prod_{j\neq i}^{t}(x_ix_j)^{s_{ij}}\ .$$

If an index k gives the contribution ≥ 1 to the sum (2.21), *i.e.* to the sum

$$\sum_{i=1}^{t}\sum_{j\neq i}^{t}s_{ij}=2r$$

then the double product

$$\prod_{j=1}^{t}\prod_{j\neq i}^{t}(x_ix_j)^{s_{ij}}$$

contains x_k as factor at least in the power 2. Hence differentiating the identity twice with respect to each x_k, $k = 1,2,\ldots,t$ and then putting all x_n equal to 1 we get the inequality

$$(2.32)\qquad 2^t N(t) \le \Big\{\prod_{k=1}^{t} \frac{\partial^2}{\partial x_k}\Big(\sum_{i=1}^{t}\sum_{j\ne i}^{t} x_i x_j\Big)^{2r}\Big\}_{x_k=1,\ k=1,2,\ldots,t}\ .$$

The right hand side, however, is at most equal to

$$(2.33)\qquad \Big\{\prod_{k=1}^{t} \frac{\partial^2}{\partial x_k}\Big(\Big(\sum_{i=1}^{t} x_i\Big)^{4r}\Big)\Big\}_{x_k = 1,\ k=1,\ldots,t} = \frac{(4r)!}{(4r-2t)!}\cdot t^{4r-2t}\ .$$

Combining (2.30), (2.32) and (2.33), we get

$$E[(T_n-\hat{T}_n)^{2r}] \le c(r)\,\|\psi'\|^{2r}\Gamma_{nr}^{2r}$$

with

$$c(r) = 2^{2r} n^{-2r}\sum_{t=1}^{2r}\binom{n}{t}\frac{(4r)!}{(4r-2t)!}\,t^{4r-2t}\cdot 2^{-t}$$

$$\Gamma_{nr}^{2r} = \frac{1}{n}\sum_{i=1}^{n}|c_i|^{2r}\ .$$

We estimate $c(r)$ exactly in the same way as we have estimated $b(r)$ in Lemma 2.1 and then obtain for $u = 2r - t$

$$c(r) \leq \sum_{u=0}^{r-1} k(u)$$

with

$$k(u) = n^{-u} \frac{(4r)!}{(2u)!(2r-u)!} (2r-u)^{2u} \cdot 2^{u} .$$

Hence

$$k(0) = \frac{(4r)!}{(2r)!} , \; k(1) < n^{-1} \cdot (2r)^{3} \frac{(4r)!}{(2r)!}$$

and for $u \geq 1$

$$\frac{k(u+1)}{k(u)} \leq \frac{4}{3} n^{-1} r^{3} \leq \frac{1}{2} \quad \text{for} \quad n^{-1} r^{3} \leq \frac{3}{8} .$$

Hence for $n^{-1} r^{3} \leq \frac{3}{8}$

$$c(r) \leq \frac{(4r)!}{(2r)!} [1 + 8n^{-1} r^{3}] .$$

Thus we have proved (2.13) and (2.14) of the lemma.

It follows by the definition of T_n that

$$S_n - T_n = \sum_{i=1}^{n} c_i \,[\xi_i - E(\xi_i)]$$

with

$$|\xi_i| \leq \frac{1}{2}(\rho_i - \rho_{ii})^2 \|\psi''\| \quad .$$

Hence

$$E[(S_n - T_n)^{2r}] \leq n^{2r-1} \sum_{i=1}^{n} c_i^{2r} E[(\xi_i - E\xi_i)^{2r}]$$

and by Lemma 2.1

$$E[(\xi_i - E(\xi_i))^{2r}] \leq 2^{2r} E[\xi_i^{2r}]$$

$$\leq \|\psi''\|^{2r} E[(\rho_i - \rho_{ii})^{4r}] \leq n^{-2r} b(2r) \|\psi''\|^{2r} \quad .$$

Thus we get (2.11)

$$E[(S_n - T_n)^{2r}] \leq b(2r)\Gamma_{nr} \; .$$

By Minkovski's inequality we obtain (2.12) from (2.10) and (2.11)

$$E^{\frac{1}{2r}}[(S_n-\hat{T}_n)^{2r}] \leq E^{\frac{1}{2r}}[(S_n-T_n)^{2r}]+E^{\frac{1}{2r}}[(T_n-\hat{T}_n)^{2r}] \; .$$

Lemma 2.3: $\hat{T}_n = \sum_{j=1}^{n} \hat{T}_n^{(j)}$ <u>with independent random variables</u>

(i) $$\hat{T}_n^{(j)} = c_j\{\psi(\rho_{jj}) - E[\psi(\rho_{jj})]\}$$

$$+ \frac{1}{n} \sum_{i\neq j}^{n} c_i[E(u(X_i-X_j) - F_j(X_i))\psi'(\rho_{ii})|X_j] \; .$$

Further

(ii) $$\sum_{j=1}^{n} [E|\hat{T}_n^{(j)}|^3] \leq 4[2\|\psi\|^3 + \|\psi'\|^3] \sum_{j=1}^{n} |c_i|^3 \; .$$

<u>Proof</u>: We get the representation (i) by (2.16). Using well-known inequalities

$$|(a+b)^3| \leq 4[|a|^3+|b|^3] \; , \; |(\sum_{i=1}^{n} a_i)^3| \leq n^2 \sum_{i=1}^{n} |a_i|^3$$

we obtain

$$E[|\hat{T}_n^{(j)}|^3] \leq 4|c_j|^3 E[|[\psi(\rho_{jj})] - E\psi(\rho_{jj})|^3]$$

$$+ \frac{4}{n} \sum_{i\neq j}^{n} |c_i|^3 \|\psi'\|^3 \; .$$

Here

$$E[|\psi(\rho_{jj}) - E[\psi(\rho_{jj})]|^3] \leq 2\|\psi\| E(\psi(\rho_{jj}) - E(\psi(\rho_{jj}))^2 \; .$$

Thus we get (ii).

3. Proofs of the theorems.

(a) Proof of Theorem 1.1 : (1.10) follows from Berry-Esséen's inequality and Lemma 2.3 and (1.11) from Lemma 2.2 (2.12).

(b) Proof of Theorem 1.2. For $h > 0$ we get

(3.1) $$P[S_n \le \hat{\delta}_n x] \le P(S_n \le \hat{\delta}_n x \text{ , } |S_n - \hat{T}_n| < h\hat{\delta}_n)$$

$$+ P[|S_n - \hat{T}_n| \ge h\delta_n] \le P[\hat{T}_n \le \hat{\delta}_n (x+h)] + P[|S_n - \hat{T}_n| \ge h\hat{\delta}_n] .$$

Applying Theorem 1.1 we get

(3.2) $$P[\hat{T}_n \le \hat{\delta}_n (x+h)] \le \phi(x+h) + 4C(2\|\psi\|^3 + \|\psi'\|^3) \cdot \sum_{i=1}^{n} |c_i|^3 \hat{\delta}_n^{-3} .$$

Here

(3.3) $$\Phi(x+h) \le \Phi(x) + h\|\Phi'(\omega)\| = \Phi(x) + \frac{h}{\sqrt{2\pi}} .$$

By Chebychev's inequality and the inequality (2.12) of Lemma 2.2 we get

(3.4) $$P[|S_n - \hat{T}_n| \ge h\hat{\delta}_n] \le d(r,\psi)\Gamma_{nr}^{2r}(h\hat{\delta}_n)^{-2r} .$$

Now we choose n such that

$$\frac{h}{\sqrt{2\pi}} = d(r,\psi)\Gamma_{nr}^{2r}(h\hat{\delta}_n)^{-2r}$$

i.e.

$$h = [(2\pi)^{\frac{1}{2}} d(r,\psi)\hat{\delta}_n^{-2r}\Gamma_{nr}^{2r}]^{\frac{1}{2r+1}} \quad . \tag{3.5}$$

It follows by Lemma 2.2 (2.12), (2.13) and (2.14) and the re remark made there in Lemma 2.2 that for $n^{-1}r^3 \leq 3/8$

$$[d(r,\psi)]^{\frac{1}{2r}} \leq C'r(\|\psi'\| + \|\psi''\|)$$

with an absolute constant C' . Then it follows by (3.4) and (3.5) that

$$\frac{h}{\sqrt{2\pi}} + d(r,\psi)\Gamma_{nr}^{2r}(h\hat{\delta}_n)^{-2r}$$

$$\leq C_2[\hat{\delta}_n^{-1}(\|\psi'\|+\|\psi''\|)r\Gamma_{nr}]^{\frac{2r}{2r+1}} \quad .$$

By (3.1) - (3.6) we get the inequality (1.12) in one direction. It follows for the other direction in the same way.

4. __Acknowledgements__. The authors would like to express their sincere appreciation to Professor I.R. Savage, the associate editor, and the referees for the critical examination of the original draft. Their constructive criticism,

detailed comments and suggestions for improvements are gratefully acknowledged.

REFERENCES

[1] Jurečkova', J. and Puri, Madan L. (1975). Order of normal approximation for rank test statistics distribution. Ann. Prob. 3 526-533.

[2] Ha'jek, J. (1962). Asymptotically most powerful rank order tests. Ann. Math. Statist. 33 1124-1147.

[3] Ha'jek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. 39 325-346.

[4] Ha'jek, J. and Sidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.

[5] Zolotarev, V.M. (1967). A sharpening of the inequality of Berry-Esseen. Z. Wahrcheinlich-keitstheorie und Verw. Gebiete. 8 332-342.

FILM

4